Trip Report

September-October 2007

Elizabeth Post

Lincoln Ventures Ltd

Attendance at conferences:

UK e-Science 2007 All Hands Meeting
10th-13th September 2007, Nottingham, UK

APAC 07 – Conference and Exhibition
9th-12th October 2007, Perth, Australia

Visits to:

National e-Science Centre and e-Science Institute, Edinburgh, UK

Open Middleware Infrastructure Institute (OMII-UK), UK

Discussions with senior people in:

Digital Curation Centre (DCC). UK

National Grid Service (NGS), UK

Visits to:

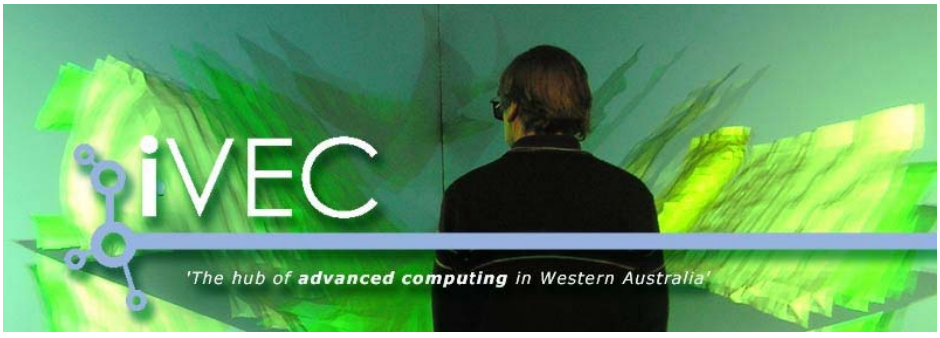
iVEC – “the hub of advanced computing in Western Australia”

Western Australian Supercomputer Program (WASP)

National e-Science Centre

e-Science Institute

omii-uk
www.omii.ac.uk



Western Australian Supercomputer Program

1 Introduction and Acknowledgements

In September/October 2007 I attended the UK All-Hands Meeting and the Australian APAC 2007 conference, and visited a number of e-science institutes or centres in both the UK and Australia, where I talked with leading people in these organisations This trip was supported by KAREN Capability Build Fund, Lincoln Ventures Ltd and Lincoln University, and I am very grateful for their support.

2 Conferences

The UK All-Hands meeting is the annual primary meeting for e-science (e-research) in the UK, and perhaps the premiere e-research conference in the world. The APAC conference is the biennial conference for the Australian Partnership for Advanced Computing which includes e-research as well as high-performance computing.

2.1 UK All-Hands Meeting 2007

The UK e-Science All Hands Meeting (AHM) provides a forum for discussion and demonstration of a broad range of e-Science projects from all disciplines.

The conference features many excellent opportunities for networking through presentations by researchers who are active in e-Science projects, poster sessions, workshops and demonstrations. The schedule also includes a number of invited Keynote speakers involved in leading Grid and e-Science activities worldwide.

This popular and well established conference is now in its sixth year and attracts delegates from many disciplines and organisations, including senior decision makers working in research and IT using advanced computing techniques and technologies. It also attracts distinguished members of the UK academic community and all eight UK Research Councils exhibit at the event.

Over 600 delegates attended this year's unique AHM event from both the UK and abroad. In 2007 this included at least 3 delegates from NZ. This is a great improvement as up until this time I have usually been the only NZ delegate at these conferences.

The full programme for this conference is available at http://www.allhands.org.uk/ and the presentations and papers can be downloaded from this site. Thus I will summarise the main points of the conference below.

2.1.1 AHM Keynote addresses

One of the main highlights of the AHM is the excellent keynote addresses. These presentations are available on the website, but I will summarise the main points below, especially for those keynotes most relevant to NZ.

Keynote 1: e-Science – the way ahead
Professor Malcolm Atkinson, UK e-Science Envoy & Director of the e-Science Institute

This address described how research goals drive the agenda, the e-Science community are guardians to ensure increasing engagement is our responsibility within existing and new disciplines, balancing organic growth with sharing and synergy. This requires investment in planning and coordination. It

noted how research using e-science, research enabling e-science and e-science infrastructure are all closely interlinked and each needs the others to exist. Also diverse applications domains, diverse enabling domains and collaboration, communication, sharing and coordination are similarly interlinked and mutually dependent. Grid technologies are transforming the way we think. It then referred to plans for future investment in cyberinfrastructure outside the UK, including the USA NSF'S *Cyberinfrastructure Vision For 21st Century Discovery* (http://www.nsf.gov/od/oci/ci_v5.pdf), the *European Roadmap on Research Infrastructures (Report 2006)* (ftp://ftp.cordis.europa.eu/pub/esfri/docs/esfri-roadmap-report-26092006_en.pdf) and the UK's proposal *Developing the UK's e-infrastructure for science and innovation* (http://www.nesc.ac.uk/documents/OSI/report.pdf). There has also been a study of *Users' Priorities for E-Infrastructure for Research* in the UK (www.nesc.ac.uk/technical_papers/uk.html). All these reports are worth studying to help us identify what needs to be done in NZ.if we are to have a place in modern e-research.

This talk then commented on recent developments in UK cyberinfrastruture with the National Grid Service (NGS) and Open Middleware Infrastructure Institute OMII-UK It also refers to a number of other e-research activities, including the opening of the Oxford e-Research centre. In 2007 the UK National e-Science Centre hosted a visit from the Swedish Research Council on "How can Science be improved by e-Science". (Perhaps it would be helpful for NZ to send a similar delegation to visit NeSC?)

The talk concluded with suggestions of the way forward in the UK for e-science, noting that the challenge is there, and it is essential to raise the level of engagement and provision for UK research.

Keynote 2: A Design of the GEO Grid: Systems of Systems federating Geospatial Data and Services
Mr Satoshi Sekiguchi: National Institute of Advanced Industrial Science and Technology, Japan

This talk describes Japan's GEO Grid. The GEO (Global Earth Observation) Grid is aiming at providing an e-Science Infrastructure for worldwide Earth Sciences communities to accelerate geosciences based on the concept that relevant data and computation are virtually integrated with a certain access control and ease-of-use interface those are enabled by a set of Grid and Web service technologies. There are many important uses for this work, including disaster prevention and mitigation, e.g. earthquakes, landslides and volcanoes, as well as studying the impact of change caused by development. This work included using workflows and data searches. This talk discussed some of the challenges faced in establishing GEO grid and an overview of how it was implemented.

Keynote 3: The Grand Challenge of Population Ageing: e-Science to the Rescue:
Professor Thomas Kirkwood: University of Newcastle, UK

This talk was very interesting as it described results that could only have been achieved by using e-science technologies, and demonstrates how these technologies are transforming the way research is done and enabling new results that have not been possible before using traditional research methods, as there is an important need to integrate a wide variety of data and hypotheses. Thus they developed Biology of Ageing e-Science Integration and Simulation (BASIS) System. The web service interface to BASIS would enable biological or medical users to run the models and modify them easily, be straightforward to accommodate changes as new hypotheses emerge, allow it to be used in ways not envisaged by the developers, permit other modellers to link new models in to the Virtual Ageing Cell, etc, and convenient to use models to carry out virtual experiments. It then gave some examples of the type of research that can be carried out using BASIS.

Finally the talk summarised the Multidisciplinary Dimensions of the Grand Challenge
- Understanding of the underlying mechanisms of biological ageing.
- Identification of ways to extend healthspan via the intrinsic malleability of the ageing process.

- Technology solutions to sustain independent living and quality of life over the lifecourse.
- Contributions to education, culture and social nclusion.
- Innovative business and contributions to economic growth.

(See EU FP6 SSA "Changing Expectations of Life" at http://ageaction.ncl.ac.uk)

Keynote 4: Medical Visualization Beyond 2D Images
Professor Anders Ynnerman: Linkoping University, Sweden

Unfortunately this talk is not available on the website. However, it described some of the significant advances in medical visualization that are now possible using e-science technologies.

Keynote 5: JISC's role in Developing a UK e-Infrastructure
Professor John Wood: Principal Designate, Faculty of Engineering, Imperial College London

This talks focuses on JISC's role in developing UK e-infrastructure. I have reported this in some detail because it is very relevant as we begin to develop e-infrastructure in NZ. It is also important as it described the extensive infrastructure being built in the UK and Europe which is far, far more than just an advanced network, and it is very important for NZ to recognize this point too.

The term "e-infrastructure" usually refers to ICT based infrastructure to support the research process. This includes:

- networks,
- access management and other "middleware" to manage the use of networked resources,
- computer facilities and peripherals including High Performance Computing (HPC),
- online content (research data, papers and journals, bibliometric data etc.).

JISC also supports researchers by:

- Providing secure, authenticated access to a wide range of on-line content and key datasets
- Providing a range of services, tools and mechanisms to enable researchers to exploit fully the value of online resources and services.
- Funding the development of institutional repositories

JISC's current activities include:

- JISC works in collaboration with research councils and other relevant organisations, particularly through the JISC Committee for the Support of Research, which I Chair.
- JISC already provides the research community with:
    - Physical network to support research – SuperJANET5, UKLight and a new dark fibre testbed for photonics research, to inform SuperJanet6
    - Access Management for researchers – the UK Access Management Federation meets the needs of most researchers, but JISC recognises the specialised requirements of Grid/e-Science community and is funding pilot projects to enable seamless inter-working between Grid and JISC's information environment
    - Advice and guidance through initiatives such as the Digital Curation Centre and the National Centre for Text Mining

JISC also supports researchers by:

- Providing secure, authenticated access to a wide range of on-line content and key datasets

- Providing a range of services, tools and mechanisms to enable researchers to exploit fully the value of online resources and services.
- Funding the development of institutional repositories

The OSI report *Developing the UK's e-infrastructure for science and innovation* (http://www.nesc.ac.uk/documents/OSI/report.pdf) provides a template for UK developments.

JISC's formal response to the report acknowledges that it has a key role to play. JISC has formed a high-level Working Group to look at wider infrastructure issues. A recent analysis of JISC and research council activities suggests that JISC and the STFC (Science and Technology Facilities Council www.scitech.ac.uk) are leading on the provision of infrastructure services for the research community, with other research councils providing more specialist support

The key recommendations of the OSI report are:

- The UK's e-infrastructure should provide researchers with:
  - Access to the systems, services, networks and resources that they need at the point that they need them
  - Facilities to discover resources easily and use them appropriately
  - Confidence in the integrity, authenticity and quality of the services and resources they use
  - Assurance that their outputs will be accessible now and in the future
  - A location-independent physical infrastructure for combining computation and information from multiple data sources
  - Advanced technologies to support collaborative research
  - The training and skills needed to exploit the services and resources available to them.
- The e-infrastructure should allow researchers to:
  - Exploit the power of advanced information technologies and applications to continuously enhance the process of research itself
  - Collaborate and communicate securely with others, across disciplines, institutions and sectors
  - Maximise the potential of advanced technologies to support innovation and experimentation
  - Share their research outputs with others and re-use them in the future
  - Engage with industry in support of wider economic goals
- The e-infrastructure must enable:
  - The growth of knowledge transfer and the development of the commercial applications of research outputs
  - Research funders to track the outputs from the research they fund
  - The protection of individuals' privacy and work, within regulatory, legal and ethical constraints
  - The protection of intellectual property and rights management
  - The preservation of digital information output as a vital part of the nation's cultural and intellectual heritage

The JISC Support of Research committee is developing a new strategy to reflect the changing needs of the research community and to address the priorities identified in the OSI report. The strategy will inform the committee's funding recommendations for 2008/09 and subsequent Academic Years. They will continue to work with the research councils to ensure that there are no gaps in provision and to eliminate overlaps This committee needs to take account of relevant international developments, including work by:

- e-IRG (e-Infrastructure Reflection Group)
- ESFRI (European Strategy Forum on Research Infrastructures)
- US initiatives to develop a 'cyberinfrastructure'

(European Strategy Forum on Research Infrastructures (ESFRI) was launched in April 2002, and brings together representatives of the 25 EU Member States and 7 Associated States, appointed by Research Ministers, and one representative of the European Commission (EC))

Research Infrastructures of pan-European relevance provide unique opportunities for world-level research and training as well as for stimulating knowledge and technology transfer, in brief for **European Capacity Building**.

The European community is planning to construct a new generation of Capability (high-performance) and Capacity Computing (high-throughput) top-level machines at an estimated cost of 550 M€.

- Scientific computing network to be set-up at European level associated with national, regional and local centres
- Different machine architectures will fulfil the requirements of different scientific domains and applications

This talk referred to the recent Fourth European Conference on Research Infrastructures, held in Germany in June 2007. The Conference looked at the best governance models, at the development of a common European strategy, and at the international dimension of Research Infrastructures. This provided valuable feedback in the launch of the EC's 7th Framework Programme and in the updating of the European roadmap of Research Infrastructures. The Conference covered all types of Research Infrastructures including the e-infrastructures and especially distributed ones.

This talk concluded by outlining what needs to be done in the future:

- Identifying how much supercomputing power the UK needs. The requirement to balance capacity and capability
- Enabling data sets from many different sources and disciplines to be mined effectively. "Just how did the scientists and engineers work together across boundaries in the construction and running of ATLAS at CERN?"
- Matching the pull of computer scientists with the needs of the academic community. Raising aspirations and integrating e-science with the development of large networks and facilities
- How to cope with massive data sets and to protect them
- How will students and teachers know something is true? – the need for strong measures to track provenance.

Keynote 6: Digital Earth: The New Digital Commons
Dr Timothy W. Foresman: University of Maryland, USA

This was a very interesting talk that identifies some of the urgent issues that must be faced internationally that are arising from global climate change.

"The planet has a life-threatening disease, marked by massive ecological degradation and rapid climate change."

"If you look at the science that describes what is happening on earth today and aren't pessimistic, you don't have the correct data." Blessed Unrest Paull Hawken 2007

The *Digital Earth* is a compelling vision for harmonizing our collective societal actions for interoperability to harness our collective human and information resources to sustain life on planet earth.

The *Digital Earth* Vision includes:

- Technology: Visualization, Data Fusion, Simulations, Modeling, Geo-location, Interoperability , Networks, SuperComputers
- Applications: Climate Change,Disasters, History & Culture, Biodiversity, Health & Safety, Social Networks, Education, Museums & Libraries

The *Digital Earth* community includes:

- Scientists, Teachers, Citizens, NGOs, Government, Industry
- China, Japan, Canada, Czech Republic, New Zealand, Australia, Germany Israel , Hungary, United Kingdom, France, Switzerland, Kenya, USA, etc,

This presentation is well worth viewing.

Keynote 7: IXI(CO): Progressing a scientific GRID project to an end-to-end solution
Dr Thomas Hartkens:Chief Technical Officer of IXICO

This talk describes the creation of an advanced application in medical imaging. IXI stands for **I**nformation e**X**traction from **I**mages. It is:

- Grid-enabled image registration and segmentation for medical research:
  - to assess scalability of Dynamic Brain Atlas
  - to extend to data from outside the brain
  - to build an application interface
- 3 year project at King's College, London and imperial College, London) funded by e-science core programme and GSK, Philips Medical Systems, Dunhill Charitable Trust

The reasons for using Grid technologies in automated medical image analysis included:

- Huge volume of image data
  - hospitals typically produce several terabytes per year
  - results of automated image analysis can be few times larger
- Computationally demanding algorithms
  - sophisticated image analysis of one single image can take several hours (or even days)

- Multiple processing steps (Workflows)
  - one solution consists often out of many single applications
- Collaboration of different institutions
  - is often required in clinical trials or research studies

One example is a Dynamic Brain Atlas, created in 2001-2002.

The IXI components include:

- algorithms wrapped as grid services
- a workflow engine and a workflow language to combine image analysis services
- a web interface to launch workflows
- grid applications to design workflows and to control the execution
- large data set of MRI brains (~550 normal volunteers, age range 20-80 years, T1, duel echo (proton density and T2 weighted) acquired at 3 different sites, publicly available to research community

Applications of IXI include: Automatic delineation of bones

A business idea for IXI is *Automated image analysis for the Drug Development Process*. There is a crisis in drug discovery and development – drug development costs are reaching levels that some manufacturers call unsustainable.. Imaging can help. "Imaging is a key technology for assessing, accelerating the development of, and guiding the use of new therapeutic options" (George Mills, FDA). "It is estimated that between $10m and $30m could be saved per drug program through the use of advanced imaging technologies which could guide go/no go decisions." (Business Insights Journal, May 2006). The customers are the pharmaceutical and biotech companies. There is an industry need for more precise measures for better, faster, cheaper decision making

2.1.2 Other activities at AHM

There were several hundred papers in various sessions. One of the exciting things about UK e-science is it ranges through everything from Aviation Industry, Cancer Research, Environmental (e.g. Climate Change), Climatology, Archaeology, Antarctic research, Psychology, Music, Neuroinformatics, etc, as well as the traditional sciences and research into e-science technologies (e.g. security, metadata management, grid interoperability, collaboration etc) themselves.

These sessions were:

- Interoperability
- Medical Applications
- Data Management
- Programming and Systems
- Scientific Applications
- Visualization
- Security
- VREs (Virtual Research Environments): Where are we now?
- Data Handling and Service Deployment

- Social Science
- Grid Middleware
- Workflows
- VREs and eResearch
- Infrastructure
- Production Grids
- SOA (Service Oriented Architectures) and Web Services

The AHM also included a number of workshops on:

- OMII-UK (Open Middleware Infrastructure Institute)
- Building Usable Systems for the Global Environment
- Visualisation tools
- Issues in Ontology Development and Use
- Grid Computing Now!
- We have to Talk about Metadata
- Text and Grid: Research Questions for the Humanities, Sciences and Industry
- Interoperability and adaptability of text mining tools

There were also birds-of a feather sessions on:

- Public Understanding of e-Science, *Organiser: David Wallom*
  This BoF was intended to show both how the tools and techniques of e-Science can be more successfully demonstrated to the wider public, including researchers that are currently not engaged in use of e-Science. Further discussion will also highlight how the use of e-Science can demonstrate to wider audiences the underlying sciences and hopefully remove some of the mystery that is associated with cutting edge research
- National Comparisons of Uptake and Sustainability of e-Science, *Chair: Alex Voss*
- UK National Grid Service User Forum, *Chair: Andrew J Richards*
- Enabling effective widespread use of e-Infrastructure, *Chair: Mike Mineter*

There was a tutorial on OGSA-DAI.

There was also an excellent poster session with 27 excellent posters on various aspects of research in both e-research technologies and also e-research applications.

There were also a number of live demonstrations of e-research applications at the various exhibition booths. These booths included all the UK research councils and a number of e-research centres such as NeSC (National e-Science Centre), JISC and various university e-research centres.

2.2 APAC 2007

http://www.apac.edu.au/apac07/

This is the fourth conference in this series, held every two years. Highlights of this were the keynote addresses and the hands-on tutorials. There was also a tour of the Western Australia Supercomputing

Program (WASP) facility at the University of Western Australia and the Interactive Virtual Environments Centre (iVEC) at UWA and ARRC Technology Park (see section 3).

2.2.1 Keynote addresses

There were a number of excellent keynote addresses highlighting what is happening in advanced computing internationally. These were all from international speakers from the USA and Europe.

Keynote 1: Integrating National e-Science Services

Dr Brian Coghlan, Trinity College, Dublin, Eire

e-Science services are evolving towards becoming an integral part of the fundamental national research infrastructures. This creates a multidimensional integration problem: technical, organisational and support. This presentation described how the three primary Irish infrastructure providers (for compute, Grid and network) are approaching the problem, with a particular emphasis on the Grid infrastructure.

Brian Coghlan (see http://www.cs.tcd.ie/coghlan/) is the leader of the Computer Architecture and Grid Research Group (CAG) in the Department of Computer Science, Trinity College Dublin, and is the instigator of and responsible for its grid activities. CAG has extensive expertise in the grid arena through the EU DataGrid, CrossGrid, EGEE-I, EGEE-II and int.eu.grid projects, and the Irish CosmoGrid, e-INIS and WebCom-G projects. It operates the Grid-Ireland Operations Centre and the EGEE Regional Operations Centre for Ireland. Dr.Coghlan is a founding Director of Grid-Ireland, Director of the Grid-Ireland OpsCentre.

Keynote 2: Powering Cyber-Research in the 21st Century

Thom H. Dunning, Jr., Director, National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign, USA

Since its creation by the National Science Foundation (NSF) and the state of Illinois in 1986, the National Center for Supercomputing Applications (NCSA) at the University of Illinois at Urbana-Champaign (UIUC) has been a leader in the development and deployment of new computing and software technologies for the scientific and engineering community. Many recent NSF reports have noted the need for a national cyberinfrastructure for science and engineering with resources, services, and tools far beyond those currently available. To meet the needs of scientists and engineers in the 21st Century, NCSA is:

- Providing cyber-resources to ensure that the most demanding scientific and engineering science and engineering problems can be solved in a timely manner.
- Creating cyberenvironments to enable scientists and engineers to take full advantage of the resources available throughout in the national cyberinfrastructure.
- Exploring innovative computing technologies to better define the pathway to petascale computing and beyond.
- Developing advanced visualization technologies to enable scientists and engineers to understand the complex systems being studied as well as inform the public of important scientific advances.

In addition, NCSA staff collaborate with university, college and high school faculty and staff to move computing technologies and computational science into the classroom. The presentation will provide an overview of NCSA's activities in each of these areas.

Keynote 3: The new HPC Center - Where the Grid and Petascale Meet
William Kramer, General Manager, Lawrence Berkeley National Laboratory. USA

HPC centers provide a robust set of services to their client base using a centrally located set of resources for large groupings of users. HPC centers most likely support large parallel workloads. Grids are known for providing a set of widely distributed services, often for "opportunistic" processing. Many times, grids are custom organized for a coherent group of users and support sequential work loads.

What is necessary for HPC centers to support a grid - or more likely - a collection of grids? What has to change and who changes? This talk examined different approaches for Grid/HPC Center integration and discussed the success stories that are helping point the way. It reviewed the degrees of integration that are necessary and examined what happens when policy and technology develop from different points of view.

Keynote 4: EGEE: Providing a Production Grid Infrastructure for Collaborative Science

Dr, Erwin Laure, CERN/EGEE, Switzerland

EGEE (Enabling Grids for eScience) operates a large scale production Grid infrastructure federating over 240 sites from 45 countries world-wide providing over 36000 CPUs and about 5 PB of disk storage to a wide variety of scientific applications. In this talk Dr Laure reviewed the challenges and successes of EGEE in building, operating, and evolving the Grid infrastructure and highlighted a few example applications. In the second half of this talk he discussed future directions of Grids in Europe, in particular how National Grid Infrastructures will pave the way to the sustainable provision of production Grids.

Keynote 5: From SRB to iRODS: Policy Virtualisation using Rule-Based Data Grids

Dr. Reagan Moore, Director, San Diego Supercomputer Center, USA

This talk described the integrated Rule-Oriented Data System (iRODS) which is a successor to the Storage Resource Broker (SRB). iRODS minimizes the labor required to manage data grids by automating the execution of management policies. Management processes are expressed through micro-services that execute standard operations at remote storage locations. Rules are applied at each storage location to control the execution of the micro-services. Persistent state information is tracked to record the outcome of each remote operation. The state information can be queried to verify that the data grid is working correctly. The iRODS system also supports periodic rules that verify assessment criteria such as trustworthiness, compliance with patient confidentiality laws, and performance goals. Through expression of management policies as rules and management processes as micro-services, it is now possible to virtualize policy, and enforce policies across multiple administrative domains.

Keynote 6: Overtaken by Events? The Red Queen Effect, Web 2.0 and the Future of the Grid

Dr. Donald F. (Rick) McMullen, Indiana University, USA

In this talk Dr McMullen discussed the relationship between current SOA approaches for building grids for e-Science and Web 2.0 approaches as exemplified by simple, composable services from Amazon, Google and Yahoo used by millions of people every day.

Grid computing to this point has been dominated by co-evolution with Enterprise computing standards. What are the potential benefits and pitfalls of adding Web 2.0 DNA to this evolutionary process? He also discussed recent work to migrate services for real-time data sources such as instruments and sensors for doing science from an SOA/Web Services world view to a more RESTful Web2.0 approach.

Keynote 7: High Performance Computing Innovation at Chevron
Mike Netzband, Chevron

Chevron Corporation, parent of Chevron, Texaco and Caltex, engages in oil and gas exploration; refining and marketing of oil, lubricants, fuels. The Energy Industry is a very capital intensive business with assets that are active for decades. It is also a very information intensive business with information that is pertinent over a wide range of time scales. Over the past fifty years, this mature paper-based industry has transitioned to a digitally enabled industry with all significant decision making processes dependent on high quality information that is digitally processed and delivered reliably and cost effectively.

High Performance Computing has played a significant role in enabling digitally-based decision making throughout the value chain for this business. From Exploration and Production through Refining and Marketing, significant computing resources are required to maintain a competitive position in a highly competitive market.

This talk described Chevron's long history of innovation with High Performance Computing which has helped maintain its leadership position in the industry and its favorable position as a partner of choice.

Keynote 8: TeraGrid: Sharing the PetaScale
Dr Dane Skow, Director, TeraGrid Grid Infrastructure Group, University of Chicago/Argonne National Laboratory

Dr Skow discussed recent experiences in TeraGrid operations and a roadmap for future developments. The TeraGrid Project is the United States National Science Foundation's CyberInfrastructure project to link together its major HPC assets. Currently there are over 250 TFlops of aggregate compute power and 20 PB of storage allocated to the US Research community through a single facility operations. 2006 was a breakthrough growth year in terms of resources, usage, integration, and community and we are enjoying in 2007 the first fruits of the substantial hardware investment program made to refresh NSF's shared SuperComputing capabilities.

2.2.2 Tutorials

The APAC conference always runs several hands-on tutorials in e- research technologies. This year they held the following workshops and tutorials:

- Australian Access Workshop (on identity and access management)
- APAC Student Forum (where students presented papers)
- DRISHTI Visualisation Tool Workshop
- Data Life Cycle Management Forum (an industry forum discussing data management issues)
- SRB (Storage Resource Broker) and iRODS (integrated Rule-Oriented Data System) workshop (data management)
- Kepler Workshop (workflows)
- Driving e-Research Collaboration across the Pacific – this workshop was an international workshop by invitation only held by Access Grid, involving participants including those from various countries (including New Zealand, Australia, UK and USA) attending the APAC conference and also leading people in the USA.
- OPeNDAP Workshop

- NIMROD Family of Tools

I attended the following tutorials and workshop:

Data Life Cycle Management Forum

This was looking at he issues data centres at large organisations, both businesses and universities, face. These included having policies and processes for data management and curation, and also addressed issues such reliability of media, such as tapes and hard disks. It is clear that despite the huge increases in data few organizations address these issues properly, and there is a huge risk in the potential for data loss, which can be very costly.

SRB and IRODS workshop

The Storage Resource Broker (SRB) has been around for quite a while and many use it, including in NZ. iRODS, integrated Rule Oriented Data Systems (http://irods.sdsc.edu), is a new open source successor to SRB and offers significantly enhanced capability. This was a hands-on workshop where we installed iRODS on your laptops and had an opportunity to experiment with setting up and executing iRODS rules to manage data. I am definitely interested in using iRODS in the future to manage shared distributed data where SRB would formerly have been chosen.

NIMROD Family of Tools tutorial

The Nimrod (www.csse.monash.edu.au/~davida/nimrod) family of tools provide mechanisms for constructing and executing large parameter sweep and search applications. The fundamental principle is that a user can supply an arbitrary unmodified application, and Nimrod can use this as the heart of a parameter sweep or search. Nimrod also allows execution of Kepler workflows.

This was an excellent tutorial where we set up parameter sweeps and executed them on supercomputing resources in Japan, the USA and Australia. It was itneresteing to see that even when jobs executing on one of these resources failed because of too many jobs on that resource, Nimrod was able to recover and automatically reallocate those jobs to another resource so that the parameter sweep completed.

3 e-Science/e-Research centres

One of the main highlights of the trip was meeting with directors of key e-infrastructure organisations in the UK, talking to them and getting their advice for NZ.

Since New Zealand is in the early stages of developing an e-research programme I specifically chose to visit e-science centres in the UK that are related to e-science infrastructure and enabling e-research in general, rather than centres concentrating on e-research projects. Two years ago, when I met with Anne Trefethen, then Director of the e-Science Core Program in the UK, and Sir John Taylor, knighted for his efforts in getting e-science started in the UK, they told me that when the UK program started they did not foresee the need for these institutes supporting infrastructure - it was only after the first year or two the need was discovered and the institutes established. Thus I thought by focusing on these institutes and what they do, New Zealand could benefit from the lessons learned in the UK.

I met with the directors of these centres in the U, except for NGS where I had discussions with other staff.

- National e-Science Centre and e-Science Institute, Edinburgh, UK
- Open Middleware Infrastructure Institute (OMII-UK), UK
- National Grid Service (NGS), UK

- Digital Curation Centre (DCC). UK

In Australia I visited two advanced computing centres in Perth as part of a tour laid on by the conference organizers.

- Australian Interactive Virtual Environments Centre (iVEC)
- Western Australian Supercomputer Program (WASP)

3.1 National e-Science Centre & e-Science Institute Edinburgh

http://www.nesc.ac.uk/
http://www.nesc.ac.uk/esi/

These two organisations are collocated in Edinburgh. The mission of the e-Science Institute (eSI) is to facilitate the e-Science community. eSI was launched in August 2001 and has excellent experience in stimulating and organising meetings for the UK's e-Science community and of hosting visitors who interact with that community. The initial programme at eSI was shaped by the pressing need to form the UK's e Science community across all disciplines, to help it develop an understanding of the challenges and available solutions, and to develop skills. That programme has been very successful.

Now the main thrust of eSI's work is changing to focus on longer-term and research centred topics. This is achieved by running themes that develop a topic over a period of six months to a year, through a series of workshops and meetings at eSI and elsewhere. The choice of themes and theme leaders is competitive, and will be recommended to the Director by the Science Advisory Board. They are already running two themes under this regime and have two more in the pipeline.

There will remain a need for a rapid response to community issues, which will continue to be met by eSI with guidance from the Programme Committee.

In particular at the National e-Science Centre and e-Science Institute, Edinburgh I was involved in a meeting with CAUDIT (Council of Australian University Directors Of Information Technology) and the NESC and NSI leading people, including Professor Malcolm Atkinson, UK e-Science Envoy and they gave us a very good briefing on the current state of e-science in the UK.

The meeting started with Alan McMeekin from Australia outlining the Australia NCRIS (National Collaborative Research Infrastructure) Strategy (http://www.dest.gov.au/sectors/research_sector/policies_issues_reviews/key_issues/ncris/) and ARIC. They said the CSIRO noted it is hard to move scientists form PCs to clusters to HPC. The Australain national data store will be each institution providing own physical stores, and then providing national strategy for federation and searching. They need people who care about it to make sure it is properly looked after.

Then Paul Bonnington outline what is happening in e-research in NZ, mentioning KAREN and describing what has been achieve in BeSTGrid so far.

Patty McMillan from Australia then spoke on the AAF (Australian Access Federation) and MAPS (Middleware Action Plan and Strategy). She mentioned how NCRIS has budgeted AU$550 million over 5 years for shared research infrastructure including $75 million for collaboration technology. The Australia Department of Trade and Industry also has budgeted AU$4.8 million for AAF. In addition ARCS (Australian Research Collaboration Service – an organisation committed to helping Australian researchers - http://www.arcs.org.au/) has also budgeted AU$20 million as part of NCRSI, for setting

up middleware at HOPC centres set up under APAC to support HPC and data, now shifting towards collaboration services. This will initially be a testbed with trial services to explore new technologies and assess when these are ready.

Peter Clarke, director of NeSC, then talked about how NGS (the National Grid Service, takes technologies and gets them ready for deployments. JISC does not have much to do with NGS.

Dave Berry, deputy director of NeSC, then talked about research and e-infrastructure development in the UK over the past 6 years. Their main success I linking people together, such as the Science and Innovation Framework which had government buy-in from DTI (Department of Trade and Industry), Treasury and Education.

The total UK e-science budget for 2001-2006 has totaled £213 million plus £100 million from JISC (mainly for infrastructure and SuperJANET). This was from the Science budget for 2003/4 to 2005/2006 and was just for staff costs, not network etc. EPSRC (Engineering and Physical Sciences Research Council) received most of the funding, and there was an industrial contribution of £25 million.

This funding has now come to an end but the Research Councils still have 3-year budget. However, although initially this funding was ring-fenced to ensure it was spent on e-science during the startup years, future funding for e-science will no longer be ring-fenced and will be expected to come out of normal funding.

Dave Berry then outlined the purposes of the various infrastructure organisations in the UK. eSI is fundamentally about progressing e-Science and NeSC is about coordination. OMII-UK's main focus to provide a repository for software developed by e-science projects, capture it and harden it so others can use it – maintaining, identifies what is useful, commissions software, e.g. GridSAM. NGS facilitates access to computational and data resources for UK researchers, and the Digital Curation Centre facilitates distributed data management.

Anna Kenway of eSI also spoke about issues in e-science in the UK. At the Global Grid Forum GGF John Taylor said if countries don't build e-infrastructure with standards etc, physicists etc will go off and hack it themselves. ESI mission to address the perceived disconnect between applications and technical experts.

The eSI also funds themes in e-science for in-depth examination for 6-12 months at £60,000 per theme, and this is proving to be a backbone to eSI activity.

Anna talked about some of the problems in uptake of e-science technologies. Interesting problems need usable solutions. A group wants to get research done, not develop middleware. On the other hand domain scientists require useable solutions. eSI helps the two groups and users to talk to each other, provide support for researchers, meetings and workshops, themes, visitors. There are difficulties in getting older researchers to use new technologies. It has been particularly difficult getting funding to pay for computer scientists to build e-science solutions. Research Councils want to fund research, not technology, and many researchers are unaware of the computing expertise needed to implement e-science technologies. NERC (Natural Environment Research Council) has been very successful in getting 49% PIs for computer scientists, just over 50% staff were software engineers etc. It seems that when the case for employing computer scientists is put to the Research Councils properly they may be prepared to fund technology to enable research, not only research.

Malcolm Atkinson said that there is no simple one-size-fits all for e-infrastructure solutions in the UK. However, there is a need to get researchers switched on fast or whole community falls behind.

There can also be problems with university ITS departments making it difficult for researchers to use e-science tools, such as web 2.0 tools, and it is important for ITS departments to be cooperative and enabling for e-science to progress.

David Fergusson talked about training and education in e-science technologies. It can't all be done in one place, and little is done at NeSC itself. However, NeSC has grown from 2 staff to 11 in 3 years. It started in 2004 and up to 2007 had given 103 courses, 62 in the UK. It supports summer schools and specific requirements for training and teaching not supported by other infrastructure. It also stimulates policy and standards to enable sharing of educational resources and materials and give students freedom to do things you don't want them doing on normal infrastructure. Much of its time is spent trying to change the culture, working with people, and this is very hard. They also provide information on experts and who knows what.

It is important that the proportion of resources to train should go up. The temptation for many organisations is to put all the money in hardware and many universities give zero training to do the job. This is a really serious problem causing deterioration of abilities of staff because they get no training and this does not maintain people's ability to do international research. The assumption is that they can use people, and replace people. However, there are really deep issues – dealing with change management, universities want people to use new methods, and yet do nothing to equip people to do that. Investment should go to training and equipping rather than hardware. There is a much better return on investment if training is given as this increases research activity and outputs.

The meeting concluded with a shared lunch.

3.2 Open Middleware Infrastructure Institute (OMII-UK)

http://www.omii.ac.uk/

OMII-UK provides software and support to enable a sustained future for the UK e-Science community and its international collaborators. Many researchers would benefit from access to emerging e-Infrastructure - such as the Grid - but are put off by software that is difficult to implement and use. OMII-UK's Software Solutions solve these problems by providing the functionality needed by researchers in an easy-to-use and easy-to-install package.

Their first three Software Solutions address the primary needs of researchers: accessing computing power, enabling the sharing of data in collaborations and automating scientific experiments with workflows. In addition to these Software Solutions, they also provide the OMII-UK Development Kit, which can be used to create researchers' own bespoke software.

They also support open-source software development by investing in community developers to produce the functionality required by their user community. They draw upon this and other software to provide a Development Kit - an easy to install and use open-source software package that provides a secure web service hosting environment, web services and the necessary tools and environments to access these services.

I met with Neil Chue Hong, director of OMII-UK, in Edinburgh, who gave me the following information. He told me how OMII and NGS (National Grid Service) have been created out of necessity to support e-science in the UK, and they work together in collaboration without overlapping. OMII mainly works in creating middleware through providing services. NGS does the deployment and running of services, OMII does the sustaining and supporting of software. OMII tries to determine best practice in integrating middleware and other services.

OMII-UK is funded primarily by EPSRC (Engineering and Physical Sciences Research Council and through GridCore from all the UK Research Councils. Currently it gets no funding from JISC, although this is expected to change to fund certain aspects of provision of infrastructure. OMII currently has about 40 people, including developers and administrators, at three sites (Southhampton, Edinburgh and Manchester).

There are two main aspects to what OMII-UK does. Firstly it takes potentially generic software from e-science projects, improves the quality, and makes it generic, to improve its long term sustainability, so others can use it in other projects and avoid reinventing the wheel. They also provide documentation and test it thoroughly to allow further development. Secondly they may be directly commissioned to provide specific software, based on surveys of what is needed. Thus they are responsive and also proactive.

Dr Hong suggested that one way to increase the uptake and use of e-research technologies is to get postgraduate students excited about them, as these people will be the next generation of academics.

3.3 National Grid Service (NGS), UK

http://www.grid-support.ac.uk/

The National Grid Service (NGS) aims to provide coherent electronic access for UK researchers to all computational and data based resources and facilities required to carry out their research, independent of resource or researcher location. The NGS's principal role is policy and high level coordination of UK Grid Operations.

More about the NGS can be found on their website.

3.4 Digital Curation Centre (DCC),UK

http://www.dcc.ac.uk/

The purpose of this centre is to provide a national focus for research and development into curation issues and to promote expertise and good practice, both national and international, for the management of all research outputs in digital format.

I met with Andrew McHugh, Audit and Advisory Services Manager, who explained the DCC to me. Essentially the DCC provides guidance, standards and advice to assist UK institutions to manage their data appropriately, and also engages in study and research relating to data curation. Andrew referred me to the paper *The Digital Curation Centre: A Vision for Digital Curation* (http://www.dcc.ac.uk/docs/DCC_Sardinia_paper_final.pdf) which explains data curation and I too refer readers to this, rather than my duplicating it here.

3.5 Australian Interactive Virtual Environments Centre (iVEC)

http://www.ivec.org/

Acting as the 'hub of advanced computing' in Western Australia, iVEC is a high performance computing (HPC) and visualisation centre with nodes in Perth's Technology Park and The University of Western Australia.

iVEC is a joint venture between Central TAFE, CSIRO, Curtin University of Technology, Murdoch University and The University of Western Australia. The Centre provides high performance computing

and advanced visualisation capabilities to Western Australia's education and research communities and local industries such as: mining and petroleum, medical training and research, architecture and construction, multimedia, education and urban planning.

http://www.pfc.org.au/twiki/pub/Main/Documents/ServicesiVEC.pdf

iVEC has been providing advanced computing services to the Western Australia research community since its inception in 2000. The iVEC partnership has received over $5.2 million of state funding to date. In May 2006, the WA Treasurer Eric Ripper announced that the State Government had set aside an additional $1.95 million a year over the next four years to continue funding iVEC into the next decade. In the business plan for this funding, iVEC has committed to delivering access to its physical infrastructure via grid technologies and e-Research services and is looking to maximise the state government investment by partnering with commonwealth funding agencies in the development of its physical and human infrastructure.

iVEC's strategy includes:

- The development of first class e-Research infrastructure in Western Australia via the installation of a petabyte scale data storage facility and a series of significant upgrades to the computational, networking and visualization facilities;
- Increasing the human capital necessary to both ensure seamless integration of iVEC's facilities into the National Grid and provide quality user support to users of the iVEC systems; and
- Integrating State key clients, including Government and Industry, into iVEC programs.

Through these steps iVEC sees demand for these services by the WA research community growing rapidly over the next five years.

The major user communities of iVEC are Astrophysics (Gravitational Wave and Radio Astronomy), Bioinformatics, Computational Biology and Biochemistry, Computational Material Science, Computational Chemistry, Computational Geoscience, Computational Physics, Genetic Epidemiology, Remote Sensing, Water Research, and eBusiness.

Many of these communities are to be supported by NCRIS Capability areas. Some notable examples include:

- Radioastronomy, where WA is one of the two shortlisted sites for the $1.6 billion Square Kilometer Array radiotelescope.
- Bioinformatics, where iVEC will invest $300k in 06/07 for the formation of an Informatics Facility.
- Computational geology, where iVEC possesses the expertise to develop the geoscience community specific interoperability standards required along with the ability to implement the necessary grid infrastructure to achieve a functioning geoscience grid.
- Remote sensing, where iVEC is now storing data from live satellite feeds that cover both the land mass of WA and the ocean off the coast. In the past, this data has all been offline but utilising the iVEC infrastructure this data can now all be accessed almost instantly.

In our tour we saw several examples of advanced visualization technologies, including 3-D videos, 3-D visualization used in industry, a Cave and others.

3.6 Western Australian Supercomputer Program (WASP)

http://www.wasp.uwa.edu.au/

The Western Australian Supercomputer Program has been established to provide an internationally competitive computational resource for Australian researchers. The Western Australian Supercomputer Program (WASP) promotes research and teaching by integrating leading edge advanced computing and visualisation for research and academic staff and students of The University of Western Australia (UWA), as well as advance disciplinary diversity, partnerships and excellence. The WASP will facilitate research and aid in educational advancement by integrating leading edge, supercomputer and scientific visualisations to individual administrative units, as well as multidisciplinary units across campus. The WASP will embrace this disciplinary diversity by creating partnerships, supporting and servicing supercomputer needs on campus. It will ensure that UWA retains superior computing and visualisation facilities. As a facility of IVEC, the WASP is one of Australia's most prestigious and internationally competitive supercomputer and scientific visualisation centres.

4 Conclusion and Recommendations

This was a very worthwhile trip and I am grateful for the financial support which made it possible.

It is very clear that e-research internationally is about very much more than just providing an advanced network. The network itself is just the very smallest part of necessary infrastructure, but to facilitate e-research, considerably more is needed in support infrastructure, such as training and education, and service provided by such UK infrastructure organizations such as the Digital Curation Centre, NGS, OMII-UK, eSI and NESC. In NZ we cannot realistically expect much progress in e-research until we invest more in such support infrastructure.

It is also clear from the content of the keynote address at both conference that in the UK, Australia, Europe and the USA there is considerable government buy-in, enthusiasm and support for e-research, and that this results in significant investment in e-research, way beyond just an advanced network. In all these countries hundreds of millions of dollars have been spent on advanced computing infrastructure (such as networks and supercomputing facilities) and also on capability building and data management, over the past 5-10 years.

If NZ is to make progress in e-research there is much we can learn from the UK, Australia, Europe and the USA and their experiences over the past 5-10 years. These include, but are not limited to, the following:

- **Funding**. It is clear that all these countries have significantly funded e-research to the extent of 100s of millions of $ (and £) over the past 5-10 years. Also, all these countries have budgeted further large sums of money for the next 5 years, much of it to support e-research infrastructure. This funding has provided very much more than just an advanced network and has also provided national supercomputing facilities, and also further support infrastructure such as NGS, DCC, OMII-UK, eSI and NESC. Unless NZ significantly increases its funding for e-research beyond just funding KAREN our researchers will increasingly be left out of international e-research projects. An advanced network is not enough, we need far more than that. Also the extremely limited Capability Building Funding is far too little to enable what is needed in NZ. The entire CBF fund over 2 years for the nation is less than what the UK allocates to just one showcase e-research project. We cannot demonstrate anything spectacular with so little funding, and researchers cannot achieve much without significant support and coordination. Also, NZ is so far behind other countries, such as Australia and the UK that we have a lot to catch up, and cannot afford any further delays through lack of funding. Computing technologies, such as advanced networks, change very rapidly, and the longer we take to start doing e-research properly, the farther we get behind.
- **Lobbying government, including ministries, ministers, right up to the Prime Minister**. The UK ex-prime minister Tony Blair was a strong advocate of e-research). In Australia ministers are also

enthusiastic about e-research and we need to get that same interest and enthusiasm in NZ. If we can generate enthusiasm at these levels it may help create more funding.

- **Increasing awareness** that e-research is about much, much more than just an advanced network, which is just the bare minimum needed. We need to provide much more support infrastructure, in training and education, middleware support, data curation support etc. it is unrealistic to expect people to do e-research with no support or funding for support at all.
- **National coordination.** All countries involved in e-research have national organizations to facilitate e-research, such as the NGS, DCC, OMII-UK, APAC etc. In NZ we desperately need national coordination to manage many aspects of e-research to avoid duplication of effort, and facilitate the uptake of technologies. We also need national coordination to enforce standards, such as metadata standards. Without such coordination we will have many problems trying to make e-research work.
- **Using computer scientists to implement e-research solutions**. It is very clear from experiences in the UK, Europe, Australia and the USA that e-research technologies are extremely complex and need computer scientists or software engineers to implement them. It is essential that we also recognize this in NZ and provide the funding for computer scientists to implement e-research solutions. If we do not do this then the solutions will not be of the necessary quality (This is why OMII-UK needed to be established in the UK.). The average researcher or postgraduate student does not have the necessary computing skills to implement their own solutions, and it is a much better use of their time for them to concentrate on their research area, and use computer scientists to develop the technology. In NZ we have not yet got to the stage where funders have recognized the need to provide funds for professional computing needed in research, and it is essential for future progress in research in NZ for this point to be recognized, and such funding included when funding research. Modern computing is far too complex to be done by amateurs. Business has recognized this for 30 years of more and no longer uses amateurs for computing, only professionals, and it is time research caught up and recognized it too.

All in all I much appreciate the opportunity to increase my awareness of e-research activities in other countries, and I really hope NZ can benefit form and take advantage of learning from what other countries have already done. We can make much more progress, faster, by building on lessons learned elsewhere, rather than reinventing the wheel ourselves.